STRUCTURAL MONOIDS
ASSOCIATED TO EQUATIONAL VARIETIES

PATRICK DEHORNOY

(Communicated by Andreas R. Blass)

Abstract. Monoids of elementary transformations are associated to certain equational varieties in order to describe the corresponding free objects; in some regular cases, these monoids reduce to groups, and the word problem for the free objects of the variety is connected with the existence of normal forms for the members of the associated group.

1. The general framework

Let \mathcal{V} be an equational variety \{\text{Co}\} and \Sigma be any set of generators. The free object of \mathcal{V} generated by \Sigma will be denoted by \mathcal{F}_\mathcal{V}(\Sigma). If \{E\} is a set of equations defining \mathcal{V}, and if M(\Sigma) denotes the free magma generated by \Sigma with respect to the signature of \mathcal{V} then \mathcal{F}_\mathcal{V}(\Sigma) is the quotient structure of M(\Sigma) under the least congruence \equiv_{\{E\}} that forces the equations in \{E\} to hold.

We assume that an infinite sequence of variables, say \text{Var}, is fixed. A term built from \text{Var} and the operators (i.e., an element of M(\text{Var})) will be called an abstract term, while the elements of M(\text{Var}) are simply called terms. A pair of abstract terms (a, b) is said to be balanced if and only if the set of variables occurring (at least once) in a is exactly the set of variables occurring (at least once) in b. A presentation \{E\} is said to be balanced if all equations in \{E\} (considered as pairs of terms) are balanced. Notice that if \{E\} is a balanced set of equations then any identity that is a consequence of \{E\} is still balanced. We shall give a description of \equiv_{\{E\}} in terms of some elementary transformations in the case that \{E\} is balanced. To get simpler notation, we shall assume in the sequel that the signature of \mathcal{V} reduces to a single binary operator, say \ast; the general case could be treated in a similar way.

If \phi is any mapping from \text{Var} into M(\Sigma), we still denote by \phi the unique morphism of M(\text{Var}) to M(\Sigma) that extends \phi, and write \sigma^\phi for the image under \phi of the abstract term \sigma; the set of all \sigma^\phi's when \phi ranges over all mappings of \text{Var} to M(\Sigma) is denoted by \text{Subst}_\Sigma(\sigma). Similar notation will occasionally be used for \varphi a mapping from \text{Var} to M(\text{Var}), or from \Sigma to M(\Sigma).

Received by the editors February 6, 1990 and, in revised form, June 3, 1991.
1991 Mathematics Subject Classification. Primary 08B20.
This work was supported in part by a CNRS grant PRC mathématiques et informatique.

©1993 American Mathematical Society
0002-9939/93 $1.00 + .25 per page

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
Definition. If \((\sigma, \tau)\) is a balanced pair of abstract terms and \(\Sigma\) is any (non-empty) set, \(\rho^{\sigma, \tau}_{\Sigma}\) is the partial mapping from \(M(\Sigma)\) into itself whose domain is \(\text{Subst}_\Sigma(\sigma)\) and that maps each term \(\sigma^\theta\) to the corresponding term \(\tau^\theta\).

Notice that \(\rho^{\sigma, \tau}_{\Sigma}\) is functional because we assumed that \((\sigma, \tau)\) is balanced, and therefore each variable occurring in \(\tau\) already occurs in \(\sigma\), so that the value of \(\sigma^\theta\) completely determines the value of \(\tau^\theta\). Clearly the image of \(\rho^{\sigma, \tau}_{\Sigma}\) is \(\text{Subst}_\Sigma(\tau)\). It is obvious that if \(T\) is the image of \(S\) under some mapping \(\rho^{\sigma, \tau}_{\Sigma}\) with \((\sigma, \tau)\) an equation in \((E)\), or more generally under the product of a finite number of such mappings, then \(S^{(E)} = T\) must hold. This, however, is not sufficient to obtain the whole congruence \(\equiv^{(E)}\), since \(\equiv^{(E)}\) has to identify not only the pairs \((\sigma^\theta, \tau^\theta)\) (and \((\tau^\theta, \sigma^\theta)\)) but also the pairs \((T_1, T_2)\) where \(T_2\) is obtained from \(T_1\) by replacing some subterm \(\sigma^\theta\) of \(T_1\) by the corresponding term \(\tau^\theta\). Therefore, we shall introduce for every equation \((\sigma, \tau)\) in \((E)\) not only the two mappings \(\rho^{\sigma, \tau}_{\Sigma}\) and \(\rho^{\tau, \sigma}_{\Sigma}\) but also a family of translated copies of these mappings indexed by the set of all possible positions where substitution can occur.

Since we assumed that the signature contains only a single binary operator, terms can be viewed as binary trees in the usual way, so that the tree \(S \cdot T\) is the binary tree whose left subtree is (the tree associated with) \(S\) and whose right subtree is (the tree associated with) \(T\). As addresses for nodes in trees, we use finite sequences of 0's and 1's, using 0 for the left direction. The set of all such sequences will be denoted by \(\text{Seq}\), and the empty sequence (address of the root for all trees) by \(\Lambda\). If \(S\) is a term and \(u\) is a member of \(\text{Seq}\), we write \(S_{/u}\) for the subterm (i.e., subtree) of \(S\) whose root is at position \(u\), if such a subterm exists. The set of all addresses \(u\) such that \(S_{/u}\) exists will be called the support of the term \(S\). Now we make the following

Definition. (i) For \(u\) in \(\text{Seq}\) and \(F\) a partial mapping from \(M(\Sigma)\) into itself, we let \(\text{tr}_u F\) be the '\(u\)-translated' copy of \(F\), i.e., the partial mapping \(G\) from \(M(\Sigma)\) into itself such that

- \(S\) is in \(\text{Dom} G\) if and only if \(S_{/u}\) exists and is in \(\text{Dom} F\);  
- in that case, \(G(S)\) is the unique term \(T\) such that \(T_{/u}\) exists and is \(F(S_{/u})\), and \(T_{/v}\) is equal to \(S_{/v}\) for every \(v\) such that \(S_{/v}\) exists and \(v\) is incompatible with \(u\) with respect to the prefix ordering on \(\text{Seq}\),

(ii) \(\mathcal{D}^{(E)}\) is the monoid generated by all mappings \(\text{tr}_u \rho^{\sigma, \tau}_{\Sigma}\) for \(u\) in \(\text{Seq}\) and \((\sigma, \tau)\) or \((\tau, \sigma)\) in \((E)\), using composition.

If we extend in the obvious way the usual definition of an action on a set to the case of partial transformations (we do not assume that the transformations are extended by the identity outside their domains), we easily get the following description of free members in the variety \(\mathcal{V}\).

Lemma 1. Assume that \((E)\) is a balanced set of equations for \(\mathcal{V}\) and \(\Sigma\) is a nonempty set; then the monoid \(\mathcal{D}^{(E)}\) acts on \(M(\Sigma)\) and \(\mathcal{V}(\Sigma)\) is the corresponding homogeneous space.

We shall see in Lemma 3 that \(\mathcal{D}^{(E)}\) is essentially the set of all identities that hold in \(\mathcal{V}\), and the point we wish to emphasize here is the fact that a natural monoid structure can be put on this set. This approach should be compared with the categorical approach that introduces groupoids and has led recently to
important developments in geometry and, in particular, in the domain of Yang-Baxter equations. (For an extended bibliography on this viewpoint, which seems to go back to Mac Lane, see [Ca].) It seems, however, that the present approach leads to different developments, such as the one introduced below concerning word problems.

The aim of this note is not to present a complete study of the monoids $\bar{\theta}_\Sigma^{(E)}$, but only to describe a few examples where they behave nicely. Before doing that, we will nevertheless establish that these monoids are rather intrinsic objects attached to the variety $\mathcal{V}$, since they are largely independent of the choices of $\Sigma$ and $(E)$. We begin with two technical auxiliary results.

**Lemma 2.** Let $\Sigma$ be a nonempty set; then the sets $\text{Subst}_\Sigma(\sigma)$ and $\text{Subst}_\Sigma(\sigma')$ are equal if and only if there exists a permutation $\pi$ of $\text{Var}$ such that $\sigma' = \sigma^\pi$.

**Proof.** Assume $\text{Subst}_\Sigma(\sigma) = \text{Subst}_\Sigma(\sigma')$. If the supports of $\sigma$ and $\sigma'$ are not equal then there exists $u$ in $\text{Seq}$ such that $u$ is in the support of $\sigma$ but not in the support of $\sigma'$ (or conversely). Let $\varphi$ be any mapping of $\text{Var}$ into $\Sigma$: the support of $\sigma^\varphi$ is the support of $\sigma'$, so that $u$ is not in the support of $\sigma^\varphi$. But for any term $S$ in $\text{Subst}_\Sigma(\sigma)$, $u$ is in the support of $S$, so $\sigma^\varphi$ cannot be in $\text{Subst}_\Sigma(\sigma')$. Hence the supports of $\sigma$ and $\sigma'$ must be equal. Now we construct a permutation $\pi$ of $\text{Var}$ such that $\sigma' = \sigma^\pi$ by successively considering the extremal points $u$ in the common support of $\sigma$ and $\sigma'$ (the addresses of the leaves in $\sigma$ and $\sigma'$) and setting $\pi(\sigma_{/u}) = \sigma'_{/u}$. An obstruction in this process occurs when treating $u$ only if one has, for some other $v$, either $\sigma_{/u} = \sigma'_{/v}$ and $\sigma'_{/u} \neq \sigma_{/v}$ or $\sigma_{/u} = \sigma_{/v}$ and $\sigma'_{/u} \neq \sigma'_{/v}$.

Assume the first case, and let $S$, $T$ be distinct members of $M(\Sigma)$. Then define $\varphi$ by $\varphi(\sigma_{/u}) = S$ and $\varphi(\sigma_{/v}) = T$ for all other variables $v$; the term $\sigma^\varphi$ is in $\text{Subst}_\Sigma(\sigma')$, but it cannot be in $\text{Subst}_\Sigma(\sigma)$, since $R_{/u} = R_{/v}$ holds for every $R$ in $\text{Subst}_\Sigma(\sigma)$, while $\sigma^{\varphi_{/u}}$ is $(\sigma'_{/u})^\varphi$, that is $S$, and $\sigma^{\varphi_{/v}}$ is $(\sigma'_{/v})^\varphi$, that is $T$. So we conclude that the construction of $\pi$ succeeds. \(\square\)

**Lemma 3.** For every nonempty $f$ in $\bar{\theta}_\Sigma^{(E)}$, there exists a balanced pair of terms $(\sigma_f, \tau_f)$ that is unique up to a permutation of the variables such that $f$ is equal to $\rho^{\sigma_f, \tau_f}$; moreover the equation $(\sigma_f, \tau_f)$ is derivable from $(E)$.

**Proof.** The uniqueness of the terms $\sigma_f$ and $\tau_f$ follows from Lemma 2. The existence is proved inductively on the length of an expression of $f$ as the product of a finite family of generators. First, if $(\sigma, \tau)$ is an equation in $(E)$, the property holds for $\rho^{\sigma, \tau}$ by definition. It holds for $\text{tr}_u \rho^{\sigma, \tau}$ as well: $\text{tr}_0 \rho^{\sigma, \tau}$ (resp. $\text{tr}_1 \rho^{\sigma, \tau}$) is $\rho^{\sigma^y, \tau^y}$ (resp. $\rho^{\sigma^y, \tau^y}$), where $y$ is any variable not occurring in $\sigma$, and the proof then goes inductively on $u$ in $\text{Seq}$. Therefore, it remains to show that the property holds for $g \circ f$ whenever it holds for $f$ and $g$. According to term unification theory (see, e.g., [HO]), there exists a term $v$ such that $\text{Subst}_\Sigma(\nu)$ is the intersection of $\text{Dom} g$ and $\text{Im} f$, i.e., of $\text{Subst}_\Sigma(\sigma_g)$ and $\text{Subst}_\Sigma(\tau_f)$. So there exist two mappings, say $\varphi$ and $\psi$, of $\text{Var}$ into $M(\text{Var})$ such that $v$ is both $\tau_f^\psi$ and $\sigma_g^\psi$. Let $\sigma$ and $\tau$ be respectively $\sigma_f^\psi$ and $\tau_g^\psi$. If $S$ is in the domain of $g \circ f$, $f(S)$ is in $\text{Dom} g$ and in $\text{Im} f$, hence in $\text{Subst}_\Sigma(\nu)$, say $f(S) = v^\sigma$. We get $f(S) = (\tau_f^\nu)^\sigma = \tau_f^\nu^\varphi$, and therefore $S = \sigma_f^\nu^\varphi = (\sigma_f^\nu)^\varphi = \sigma^\varphi$, so $S$ is in $\text{Subst}_\Sigma(\sigma)$. Moreover, we get also $f(S) = (\sigma_g^\psi)^\varphi = \sigma_g^\psi$, and therefore $g(f(S)) = \tau_g^\sigma = (\tau_g^\sigma)^\varphi = \tau^\varphi$, and this proves that $g \circ f$ is exactly $\rho^{\sigma, \tau}$. Next, we notice that the pair $(\sigma, \tau)$
is balanced. Indeed, assume that \( x \) occurs in \( \sigma : x \) occurs in some \( y^\varphi \) for some \( y \) occurring in \( \sigma_f \), hence in \( \tau_f \), so \( x \) occurs in \( \tau_f \), i.e., \( \sigma_g^\varphi \). Hence \( x \) occurs in some \( z^\varphi \) for some \( z \) occurring in \( \sigma_g \), hence in \( \tau_g \), so \( x \) occurs in \( \tau_g^\varphi \), i.e., \( \tau \). The converse argument also works, so we are done. Finally, the equation \( \sigma = \tau \), i.e., \( \sigma_f^\varphi = \tau_g^\varphi \), follows (since \( \tau_f^\varphi \) and \( \sigma_g^\varphi \) are equal) from \( \sigma_f^\varphi = \tau_f^\varphi \) and \( \sigma_g^\varphi = \tau_g^\varphi \), hence from \( \sigma_f = \tau_f \) and \( \sigma_g = \tau_g \), and finally from \( (E) \) by the induction hypothesis. \( \square \)

**Definition.** (i) We say that a term \( \sigma \) (resp. a pair of terms \( (\sigma, \tau) \)) is a *strict instance* of another term \( \sigma' \) (resp. another pair \( (\sigma', \tau') \)) if and only if \( \sigma = \sigma^\varphi \) holds (resp. \( \sigma = \sigma^\varphi \) and \( \tau = \tau^\varphi \) hold) for some \( \varphi \) such that \( x^\varphi \) is not a variable for some variable \( x \) occurring in \( \sigma \), or \( x^\varphi = y^\varphi \) holds for two distinct variables \( x, y \) occurring in \( \sigma' \).

(ii) we say that a set of equations \( (E) \) is *minimal* if and only if no equation in \( (E) \) is a strict instance of some consequence of \( (E) \).

For instance, the set \( \{ x = x \ast y, (x \ast y) \ast (x \ast y) = x \ast y \} \) is not minimal, because \( x \ast x = x \) is a consequence of these equations and the second equation is a strict instance of \( x \ast x = x \). (The preceding set is not balanced, but balanced nonminimal sets of equations can be easily constructed.) Minimal sets of equations always exist for a given variety. Starting from any set of equations and substituting for a ‘nonminimal’ equation another equation of which this nonminimal one is a strict instance gives a new set of equations for the same variety. Then iterating this process must give within a finite number of steps a minimal set of equations since the cumulated weights of the equations must decrease when the process is performed, where the weight of an equation is defined to be the difference of its length and of the number of distinct variables occurring in it. Notice that this reduction process does not destroy the property of being balanced.

The invariance result we can get in the general case is

**Proposition 4.** Assume that \( \Sigma, \Sigma' \) are nonempty sets and that \( (E), (E') \) are minimal sets of equations defining the same variety \( \mathcal{V} \); then the monoids \( \vartheta^E_\Sigma \) and \( \vartheta^E_\Sigma' \) are isomorphic.

**Proof.** We begin with the independence with respect to \( \Sigma \). Equipotent \( \Sigma \)'s clearly produce isomorphic \( \vartheta^E_\Sigma \)'s, so it suffices to consider the case that \( \Sigma' \) is a subset of \( \Sigma \). For \( f \) in \( \vartheta^E_\Sigma \), \( f \mid M(\Sigma') \) is in \( \vartheta^E_{\Sigma'} \) because \( \rho^\Sigma_{\sigma, \tau} \) is the restriction of \( \rho^\Sigma_{\sigma, \tau} \) to \( M(\Sigma') \). Thus the mapping \( f \mapsto f \mid M(\Sigma') \) gives an epimorphism of \( \vartheta^E_\Sigma \) onto \( \vartheta^E_{\Sigma'} \). We show that this epimorphism is one-one. Consider any \( f \) and \( g \) in \( M(\Sigma) \), and assume \( f \mid M(\Sigma') = g \mid M(\Sigma') \). By Lemma 3, there exist terms \( \sigma_f, \sigma_g, \tau_f, \tau_g \) such that \( f \) and \( g \) are respectively \( \rho^\Sigma_{\sigma, \tau} \) and \( \rho^\Sigma_{\sigma, \tau} \). Then the domain of \( f \mid M(\Sigma') \), that is \( \text{Subst}_{\Sigma'}(\sigma_f) \), is equal to the domain of \( g \mid M(\Sigma') \), that is \( \text{Subst}_{\Sigma'}(\sigma_g) \). By Lemma 2, there exists a permutation \( \pi \) of \( \text{Var} \) such that \( \sigma_g = \sigma_f^\pi \), and therefore \( \text{Subst}_{\Sigma}(\sigma_f) \) and \( \text{Subst}_{\Sigma}(\sigma_g) \), which are \( \text{Dom} f \) and \( \text{Dom} g \), are equal. Using the same argument, we prove that there exists a permutation \( \pi' \) of \( \text{Var} \) such that \( \tau_g \) is \( \tau_f \). \( \square \)
common support such that \( \tau_{f/\nu} \) and \( \tau_{g/\nu} \) are different variables. Let \( S \), \( T \) be two different terms in \( M(\Sigma') \), and define \( \varphi \) by \( \varphi(\tau_{f/\nu}) = S \) and \( \varphi(y) = T \) for all other variables \( y \). Then \( \sigma_{g,\varphi} \) is in the domain of \( f \) and \( g \), and we get
\[
\begin{align*}
f(\sigma_{g,\varphi})/\nu &= f(\tau_{f,\varphi})/\nu = (\tau_{f,\varphi})/\nu = (\tau_{f/\nu})^\varphi = S, \\
g(\sigma_{g,\varphi})/\nu &= (\tau_{g,\varphi})/\nu = (\tau_{g/\nu})^\varphi = T,
\end{align*}
\]
contradicting the hypothesis \( f \upharpoonright M(\Sigma') = g \upharpoonright M(\Sigma') \).

We turn to the independence with respect to \( (E) \). Assume that \( (E') \) and \( (E) \) are minimal sets of equations for the same variety and that \( \Sigma \) is infinite. Let \( (\sigma', \tau') \) be an equation in \( (E') \). Fix a one-one mapping \( \varphi \) of \( \text{Var} \) into \( \Sigma \). Then \( \sigma^{\varphi} \) is in the domain of \( \rho_{\Sigma}^{\sigma', \tau'} \) and \( \rho_{\Sigma}^{\sigma', \tau'}(\sigma^{\varphi}) \) is \( \tau^{\varphi} \). Certainly \( \sigma^{\varphi} \equiv (E')^{\tau^{\varphi}} \) holds, and so does \( \sigma^{\varphi} \equiv (E)^{\tau^{\varphi}} \). Hence, by Lemma 1, there exists \( f \) in \( \delta(\Sigma) \) such that \( f \) maps \( \sigma^{\varphi} \) to \( \tau^{\varphi} \). Now any other term in \( \text{Dom} \rho_{\Sigma}^{\sigma', \tau'} \) can be written as \( (\sigma^{\varphi})^\psi \) for some \( \psi \) (because of the particular choice of \( \varphi \)), so that \( \rho_{\Sigma}^{\sigma', \tau'}((\sigma^{\varphi})^\psi) \) is \( (\tau^{\varphi})^\psi \), i.e., \( f((\sigma^{\varphi})^\psi) \). This shows that \( \rho_{\Sigma}^{\sigma', \tau'} \) and \( f \) coincide on the domain of \( \rho_{\Sigma}^{\sigma', \tau'} \). By Lemma 3, there exist \( \sigma \) and \( \tau \) such that \( f = \rho_{\Sigma}^{\sigma, \tau} \). If \( f \) is not equal to \( \rho_{\Sigma}^{\sigma, \tau} \), the term \( \sigma \) is not in \( \text{Subst}_{\Sigma}(\sigma') \), and therefore \( (\sigma', \tau') \) is a strict instance of \( (\sigma, \tau) \). Since the identity \( \sigma = \tau \) is a consequence of \( (E) \), hence of \( (E') \), this contradicts the minimality of \( (E') \). So we conclude that \( \rho_{\Sigma}^{\sigma', \tau'} \) is in \( \delta_{\Sigma}(E') \), and it follows that \( \delta_{\Sigma}(E') \) is equal to \( \delta_{\Sigma}(E) \). □

Because of the result above, we shall from now on omit the reference to a particular set \( \Sigma \) or to a particular minimal set of equations \( (E) \) for the variety \( \mathcal{V} \), thus just writing \( \delta_{\mathcal{V}} \) for "the" monoid associated to \( \mathcal{V} \) (that exists provided that \( \mathcal{V} \) has at least one balanced presentation). Then \( \delta_{\mathcal{V}} \) is essentially the family of identities that hold in \( \mathcal{V} \): for every \( f \) in \( \delta_{\mathcal{V}} \), the associated pair \( (\sigma_f, \tau_f) \) is an identity that holds in \( \mathcal{V} \); conversely, the argument above shows that any identity that holds in \( \mathcal{V} \) is an instance of a pair \( (\sigma_f, \tau_f) \) for some \( f \) in \( \delta_{\mathcal{V}} \). The additional structure given by the monoid operation corresponds to the relationships between the identities.

2. Transformation of \( \delta_{\mathcal{V}} \) to a Group

Inverses "nearly" exist in the monoid \( \delta_{\mathcal{V}} \), since the product of \( \rho_{\sigma, \tau} \) and \( \rho_{\tau, \sigma} \) is the identity of the domain of \( \rho_{\sigma, \tau} \). Therefore, if all partial identities of the form \( \text{id}_A \) with \( A \) being the domain (or the image) of some \( k \) in \( \delta_{\mathcal{V}} \) are identified, the corresponding quotient will be a group, and, in good cases, this group will still act on the free magmas \( M(\Sigma) \). We give simple examples where such a situation occurs.

**Proposition 1.** (i) For \( f, g \) in \( \delta_{\mathcal{V}} \), write \( f \approx g \) if and only if \( f \circ k = g \circ k \) holds for some \( k \) in \( \delta_{\mathcal{V}} \); then \( \approx \) is a congruence on \( \delta_{\mathcal{V}} \) and the quotient monoid is a group \( G_{\mathcal{V}} \).

(ii) If the implication
\[
(\#) \quad f \upharpoonright \text{Dom} \ k = \text{id}_{\text{Dom} \ k} \Rightarrow f = \text{id}_{\text{Dom} \ f}
\]
holds for every \( f, k \) in \( \delta_{\mathcal{V}} \), then the group \( G_{\mathcal{V}} \) acts on \( M(\Sigma) \) and \( \mathcal{V}(\Sigma) \) is the corresponding homogeneous space.
Proof. (i) For \( f \) in \( \mathcal{D}^\mathcal{V} \), say \( f = \rho^{\sigma, \tau} \), we let \( \bar{f} \) be the corresponding \( \rho^{\tau, \sigma} \). Assume \( f \approx g \) and \( g \approx h \): by definition, there exist \( k \) and \( k' \) in \( \mathcal{D}^\mathcal{V} \) satisfying \( f o k = g o k \) and \( g o k' = h o k' \). The first equality implies \( f o k o k' = g o k o k' \) and the second implies \( g o k o k' = h o k o k' \), so that \( f \approx h \) holds. Indeed, for every \( S \) in Dom\((g o k o k')(S)\), \( k o k'(S) \) is equal to \( k'(S) \), and therefore \( g o k o k'(S) \) is equal to \( h o k o k'(S) \). Moreover, \( f o k = g o k \) implies \( h o f o k = h o g o k \) for every \( h \), i.e., \( f \approx g \) implies \( h o f \approx h o g \). And \( f o k = g o k \) implies \( f o h o k = g o h o k \) (because \( h o h o k(S) \) is equal to \( k(S) \) for every \( S \) in Dom\((f o h o k)(S)\)), so \( f \approx g \) implies \( f o h \approx g o h \) and \( \approx \) is a congruence. We claim that the quotient \( \mathcal{D}^\mathcal{V} / \approx \) is a group: indeed, for every \( f \) in \( \mathcal{D}^\mathcal{V} \), one has \( f o f o f = f \) and \( f o f o f = f \), which shows \[ f o f \approx \text{id} \quad \text{and} \quad f o f \approx \text{id}. \]

(ii) Assume \((\#)\). Then \( f \approx g \) implies that \( f \) and \( g \) agree on the intersection of their domains. Assume \( f o k = g o k \); then \( f o k o k o g \) is the identity on its domain. If \( S \) is in this domain, \( k o k o g(S) \) is equal to \( g(S) \), so \( f o g(S) \) is \( S \). Hence \( f o g \) is the identity on the domain of \( f o k o k o g \), and therefore, using \((\#)\), on its domain. This means that for \( S \) in Dom\(f \cap \text{Dom} g\), \( g(S) \) is equal to \( f o g \circ g(S) \), i.e., \( f(S) \). It follows that writing \([f]\) for the class of \( f \), we get a well-defined operation of \( G^\mathcal{V} \) on \( M(\Sigma) \) when defining \([f](S)\) to be \( T \) if and only if \( g(S) = T \) holds for some \( g \) in \([f]\), and the associated homogeneous set is still \( f^\mathcal{V}(S) \) since the orbits under this action are the same as the orbits under the action of \( \mathcal{D}^\mathcal{V} \). □

There are simple cases where condition \((\#)\), and therefore the result of the above proposition, hold.

Lemma 2. Say that a term \( \sigma \) is linear if and only if each variable occurring in \( \sigma \) occurs exactly once; if \( \mathcal{V} \) has a presentation consisting of balanced pairs of linear terms, then \( \mathcal{D}^\mathcal{V} \) satisfies \((\#)\).

Proof. We notice that (with the assumptions of the present lemma and the notations of Lemma 1.3), the associated terms \( \sigma_f \) and \( \tau_f \) are linear for any \( f \) in \( \mathcal{D}^\mathcal{V} \). It suffices to prove the property for \( g o f \) whenever it holds for \( f \) and \( g \). Using the notation of the proof of Lemma 1.3, we notice that the intermediate term \( v \) must be linear, so that the terms \( y^{\phi_r} \) for \( y \) occurring in \( \sigma_f \) are linear and have mutually distinct variables, and \( \sigma \) must be linear. The argument is similar for \( \tau \). Now, if \( \sigma \) is a linear term, \( \text{Subst}_\Sigma(\sigma) \) exactly consists of the terms whose support includes the support of \( \sigma \). For every \( a \) in \( \Sigma \), and every integer \( n \), let \( K^n_a \) be the “complete binary tree” with height \( n \) and all leaves labelled \( a \) (\( K^0_a = a \), and \( K^{n+1}_a = K^n_a * K^n_a \)), and let \( \varphi_\sigma \) be the mapping from \( \Sigma \) to \( M(\Sigma) \) that maps \( a \) to \( K^n_a \), where \( n \) is the height of \( \sigma \): then, for each term \( S \) in \( M(\Sigma) \), the term \( S^{\varphi_\sigma} \) is in \( \text{Subst}_\Sigma(\sigma) \). Now, in order to establish \((\#)\), assume that \( f \) is the identity on the domain of \( k \), and let \( S \) be any element of Dom\(f\): there exists a linear term \( \sigma \) such that \( \text{Dom} k = \text{Subst}_\Sigma(\sigma) \). Then \( S^{\varphi_\sigma} \) is in \( \text{Dom} k \), so \( f(S^{\varphi_\sigma}) \) is equal to \( S^{\varphi_\sigma} \); now \( f(S^{\varphi_\sigma}) \) is \( f(S)^{\varphi_\sigma} \) (because \( S \) is in \( \text{Dom} f \)), and we get \( f(S) = S \) since the mapping \( \varphi_\sigma \) is one-one on \( M(\Sigma) \). So \( f \) is the identity on \( \text{Dom} f \). □
3. The case of commutativity

We now turn to an effective description of the group $G^\Sigma$ in some simple cases. In the forthcoming examples, we consider presentations $(E)$ that are reduced to a single equation, say $(\sigma, \tau)$. Then there are only two elementary transformations $\rho^{\sigma, \tau}$ and $\rho^{\tau, \sigma}$ to be considered, altogether with their translated copies. We shall use simplified notation: for $u$ in Seq, the mapping $tr_{u}^{\rho^{\sigma, \tau}}$ will be denoted by $u^+$ and the mapping $tr_{u}^{\rho^{\tau, \sigma}}$ will be denoted by $u^-$. Also, we use instead of composition the opposite law on $\mathfrak{G}$, writing $fg$ for $g \circ f$, and in order to get consistent notation, writing $S^f$ for the image of $S$ under $f$.

Example. Assume that the defining equation is $x \ast y = y \ast (x \ast x)$, that $\Sigma$ is $\{a, b, c\}$, and $S$ is the term $a \ast (b \ast b)$ in $M(\Sigma)$. Then $S$ is (in particular) in the domain of $\Lambda^+ \text{, } \Lambda^- \text{, and } 1^+ \text{, and the corresponding images are easily computed as in Figure 1.}$

Our first detailed example will be the variety, say $\mathfrak{Com}$, formed by commutative magmas: $\mathfrak{Com}$ is defined by the equation $x \ast y = y \ast x$.

This equation is balanced (and minimal), and the terms $x \ast y$ and $y \ast x$ are linear, so that the preceding background applies and we get a group $G^{\text{Com}}$. We shall establish now a more precise description of this group. With the previous notation, the generator $\Lambda^+ \text{ of } \mathfrak{Com}$ is the involutory transposition operator $\Lambda^+ : S \ast T \mapsto T \ast S$.

Let $K_n$ denote a fixed complete binary tree with height $n$ wearing distinct leaves; set

$$G^{\text{Com}}_n := \{[f] \in G^{\text{Com}} \mid K_n \in \text{Dom}[f]\}.$$ 

For every $[f]$ in $G^{\text{Com}}_n$, the image of $K_n$ under $[f]$ is still a complete binary tree, so $[f]$ is determined by only the ordering of the $2^n$ leaves of $[f](K_n)$. Therefore $G^{\text{Com}}_n$ is a subgroup of the symmetric group $S_{2^n}$. Moreover (assuming that the leaves of $K_n$ are labelled $1, \ldots, 2^n$), any member of $G^{\text{Com}}_n$ either leaves both subintervals $1 \cdots 2^{n-1}$ and $(2^{n-1} + 1) \cdots 2^n$ globally invariant or it switches them, according to whether $\Lambda^+$ occurs an even or an odd number of times in any writing of this member. It follows by induction that every member of $G^{\text{Com}}_n$ has a unique expression as $[\Lambda^+]^e_0 [0^+]^e_0 [1^+]^e_1 [00^+]^e_2 \cdots$ with $e_u = 0$ or 1 (under the convention that $[f]^0$ is the identity). So we can state
Proposition 1. (i) The group $G_{n}^{\circ m}$ is the directed union of all $G_{n}^{\circ m}$'s, and $G_{n}^{\circ m}$ is a subgroup of $\mathfrak{S}_{2n}$ with $2^{2n-1}$ members.

(ii) Every member of $G_{n}^{\circ m}$ has a unique expression as a product

$$\prod_{u \in \text{Seq}} [u^{+}]^{e_{u}}, \quad e_{u} = 0 \text{ or } 1,$$

where all but a finite number of $e_{u}$ are equal to 0, and Seq is enumerated using the linear ordering $<$ that extends both the prefix ordering and the “left-right” ordering.

For instance $G_{1}^{\circ m}$ is $\mathbb{Z}/2\mathbb{Z}$, $G_{2}^{\circ m}$ is an 8 elements subgroup of $64$, namely the dihedral group $D_{4}$. Finally notice that for every $u$ with length $p$ in Seq, $u^{-}$ is a conjugate of $(1^{p}^{+})^{-}$, so that $G_{n}^{\circ m}$ is generated by the family of all $[(1^{p}^{+})^{+}]$'s. This allows us to compute (the normal form of) the product of any two members of $G_{n}^{\circ m}$ given in normal form; the rules to be applied are, for any $u$, $v$, $w$

(i) $[u0v^{+}][u1w^{+}] = [u1w^{+}][u0v^{+}]$;

(ii) $[u^{+}][u0v^{+}] = [u1v^{+}][u^{+}]$;

(iii) $[u^{+}][u1v^{+}] = [u0v^{+}][u^{+}]$;

(iv) $[u^{+}][u^{+}] = [\text{id}]$.

These relations constitute a presentation of the group $G_{n}^{\circ m}$ from its generators $[u^{+}]$'s.

4. The case of associativity

Let us consider now the variety $\mho \mathcal{G}$ formed by associative magmas (i.e., by semigroups). $\mho \mathcal{G}$ is defined by the equation

$$x \ast (y \ast z) = (x \ast y) \ast z.$$  

This equation is balanced (and minimal), and the terms $x \ast (y \ast z)$ and $(x \ast y) \ast z$ are linear, so that the background of §2 applies again, and we get a group $G^{\circ \mathcal{G}}$.

In the present case, the elementary generator $[A^{+}]$ is the class of

$$A^{+} : S * (T * U) \equiv (S * T) * U.$$  

We shall discuss the question of getting normal forms for the members of the group $G^{\circ \mathcal{G}}$. This, however, is not obvious for 'nontrivial' relations hold between the generators $[u^{+}]$ of this group, for instance

$$[A^{+}][A^{+}] = [1^{+}][A^{+}][0^{+}] .$$

This relation expresses two different ways for deriving the identity $x(yzt) = ((xy)z)t$ from associativity, and exactly corresponds to the pentagonal diagram mentioned in [Ca].

We shall get normal forms in $G^{\circ \mathcal{G}}$ by constructing an algorithm that solves the word problem for $\mathcal{G}(\Sigma)$ (i.e., $\Sigma^{+}$). As quoted in Lemma 1.1, two terms, say $S$, $T$, in $M(\Sigma)$ represent the same member of $\mathcal{G}(\Sigma)$ if and only if there exists some member $[f]$ of $G^{\circ \mathcal{G}}$ mapping $S$ to $T$. So if we are able to produce, starting from an arbitrary pair $(S, T)$, a distinguished member $f$ in $G^{\circ \mathcal{G}}$ mapping $S$ to $T$ (if it exists), we have a good chance to have exhibited in this way a canonical form for the members of $G^{\circ \mathcal{G}}$. This approach works in the present case.
It will be convenient in the sequel to use right Polish notation for terms instead of the algebraic notation. Terms in \( M(\Sigma) \) will therefore be considered as finite sequences of members of \( \Sigma \cup \{\ast\} \), i.e., as mappings of an integer interval \( 1, \ldots, N \) to \( \Sigma \cup \{\ast\} \). The following is a standard property of the terms written in right Polish notation.

**Lemma 1.** Define for \( S \) a finite sequence on \( \Sigma \cup \{\ast\} \) a mapping \( \hat{S} : \{0\} \cup \text{Dom} S \rightarrow \mathbb{Z} \) by

\[
\hat{S}(n) := \begin{cases} 
-1, & \text{if } n = 0; \\
\hat{S}(n - 1) + 1, & \text{if } n \geq 1 \text{ and } S(n) \text{ is in } \Sigma; \\
\hat{S}(n - 1) - 1, & \text{if } n \geq 1 \text{ and } S(n) \text{ is } \ast.
\end{cases}
\]

Then \( S \) is in \( M(\Sigma) \) if and only if \( \hat{S} \) takes only nonnegative values on \( \text{Dom} S \) and moreover the last value of \( \hat{S} \) is 0.

**Definition.** (i) When \( S, T \) are distinct terms in \( M(\Sigma) \), the discrepancy between \( S \) and \( T \), written as \( \delta(S, T) \), is the least integer \( k \) such that \( S(k) \) is not equal to \( T(k) \)

(ii) For \( p \) an integer and \( u \) in \( \text{Seq} \), define \( u^{(p)} \) in \( \mathcal{O}_{\mathcal{P}} \) by

\[
u^{(p)} := \begin{cases} 
i, & \text{if } p = 0; \\
u^+(u_0)^+ \cdots (u_0^{p-1})^+, & \text{if } p > 0; \\
u_0^{-(p-1)} \cdots u_0^- u^-, & \text{if } p < 0.
\end{cases}
\]

The point for getting a uniqueness result in \( \mathcal{O}_{\mathcal{P}} \) is

**Lemma 2.** Assume that \( S \) is in \( M(\Sigma) \) and \( \hat{S}(n) \) is at least 2; then there exists a unique transformation \( u^{(p)} \) such that \( \delta(S, S^{u^{(p)}}) \) is equal to \( n \).

Getting the term \( S^{u^{(p)}} \) above from \( S \) is fairly easy. It suffices to switch in \( S \) the subword \( S_0 \) and the symbol \( \ast \) that follows it, where \( S_0 \) is the subword of \( S \) made by the symbols with indices \( n + 1, \ldots, m - 1 \) and \( m \) is the least integer greater than \( n \) such that \( S(m) \) is \( \ast \) and \( \hat{S}(m) \) is \( \hat{S}(n) - 2 \). The corresponding values of \( u \) and \( p \) are determined by the fact that the 'tree address' in \( S \) of the symbol having index \( n \) is \( u_1 10^{p-1} 10^q \) for some integer \( q \).

Iterating this process gives rise to the desired algorithm. Start with two terms, say \( S \) and \( T \), of \( M(\Sigma) \); if they are not equal, let \( n \) be the discrepancy \( \delta(S, T) \), three cases may occur:

(i) If \( S(n) \) and \( T(n) \) are distinct members of \( \Sigma \) or if \( S(n) \) is defined while \( T(n) \) is not (i.e., if \( T \) is a strict prefix of \( S \)) or conversely, then quit the algorithm: \( S \) and \( T \) are not equivalent.

(ii) If \( S(n) \) is in \( \Sigma \) while \( T(n) \) is \( \ast \), then replace \( S \) by \( S^{u^{(p)}} \) where \( u^{(p)} \) is the unique transformation quoted in Lemma 2, and perform the algorithm again (the hypotheses of Lemma 2 are fulfilled for \( \hat{T}(n) \) is at least 0 so \( \hat{S}(n) \), i.e., \( \hat{T}(n) + 2 \), is at least 2).

(iii) If \( S(n) \) is \( \ast \) while \( T(n) \) is in \( \Sigma \), then proceed as in (ii) mutatis mutandis.

The algorithm stops when terms are equal.
Example. Consider the following terms:

\[ S_0 := ab \ast cde \ast \ast, \quad T_0 := abc \ast d \ast e \ast \ast \]

(that would be respectively written as \((a \ast b) \ast (c \ast (d \ast e))\) and \(a \ast ((b \ast c) \ast d) \ast e)\) in algebraic notation). The value of \(\delta(S_0, T_0)\) is 3, and one first has to transform \(T_0\) according to rule (iii). The integer \(\m\) in the proof of the lemma is here 9, and the transformation to be performed is \(\Lambda^{(3)}\) (i.e., the product \(\Lambda^+0^00^+\)) for the ‘tree address’ of 3 in \(S\) is 1001. Therefore after one step one obtains the following terms:

\[ S_1 := ab \ast cde \ast \ast, \quad T_1 := ab \ast c \ast d \ast e \ast \ast \]

The discrepancy is now 5, and \(\Lambda^+\) has to be applied to \(S_1\) since the ‘tree address’ of 5 in \(S_1\) is 110. After this second step one gets:

\[ S_2 := ab \ast c \ast de \ast \ast, \quad T_2 := ab \ast c \ast d \ast e \ast \ast \]

The discrepancy is 6, and \(\Lambda^+\) has to be applied to \(S_2\) since the ‘tree address’ of 6 in \(S_2\) is 11. In this way one finally gets equal trees, and the algorithm succeeded.

It should be clear that this algorithm is correct and that it stops eventually because at each step the discrepancy is shifted to the right while the lengths of the words do not increase. So the conclusion is that given any two equivalent terms \(S\) and \(T\) we construct two finite sequences of elementary transformations, say \((u_1^{(p_1)}, \ldots, u_r^{(p_r)})\) and \((v_1^{(q_1)}, \ldots, v_s^{(q_s)})\), such that \(Su_1^{(p_1)} \cdots u_r^{(p_r)} = Tv_1^{(q_1)} \cdots v_s^{(q_s)}\) are equal. Moreover, these sequences enjoy the following properties: if \(R\) is any term belonging to the intersection of the images of \(\prod_{i=1}^r u_i^{(p_i)}\) and \(\prod_{j=1}^s v_j^{(q_j)}\), and if \(P_i\) (resp. \(Q_j\)) denotes \(R^{u_1^{(p_1)} \cdots u_i^{(p_i)}-1} \cdots u_r^{(p_r)}\) (resp. \(R^{v_1^{(q_1)} \cdots v_j^{(q_j)}-1} \cdots v_s^{(q_s)}\)), then one has

(i) \(\delta(P_i, P_{i+1}) < \delta(P_{i+1}, P_{i+2})\) for \(i < r - 1\),
(ii) \(\delta(Q_j, Q_{j+1}) < \delta(Q_{j+1}, Q_{j+2})\) for \(j < s - 1\),
(iii) \(\delta(P_i, P_{i+1}) \neq \delta(Q_j, Q_{j+1})\) for \(i < r\) and \(j < s\).

Call such pairs of finite sequences \emph{normal pairs}. One can easily give a combinatorial characterization of the sequences that satisfy condition (i) (or (ii)) in the definition of normal pairs. For denote by \(\leq\) the linear ordering on \(\text{Seq}\) such that \(u \leq v\) holds if and only if either \(v\) is a prefix of \(u\) or \(u = w0u'\) and \(v = w1v'\) for some \(w, u', v'\); then the sequence \((u_1^{(p_1)}, \ldots, u_r^{(p_r)})\) satisfies condition (i) of normality if and only if \(u_i0^{(p_i)} \leq u_{i+1}0^{(p_i+1)}\) for \(i < r - 1\).

Then we have the following unique normal form result for the group \(G^{\varphi \varphi}\):

**Proposition 3.** Every element of \(G^{\varphi \varphi}\) has a unique expression as \(\prod_{i=1}^r[u_i^{(p_i)}] \times \prod_{j=1}^s[v_j^{(-q_j)}]\) with \(((u_1^{(p_1)}, \ldots, u_r^{(p_r)}), (v_1^{(q_1)}, \ldots, v_s^{(q_s)}))\) a normal pair.

Using this normal form result, one could rather easily determine a presentation of \(G^{\varphi \varphi}\) from its generators \(\{u^+\}\). Notice that the normal form result above for the elements of \(G^{\varphi \varphi}\) is quite effective: starting from an arbitrary expression of an element \(f\) as a product of generators, we get the normal form by applying the algorithm to any pair \((S, S')\). (Constructing a term \(S\) in the
domain of \( f \) is easy since the only obstruction for a given term to being in the domain of \( f \) is having a too small support, so that we can get the desired term \( f \) by repeated enlargements of the support.)

**Example.** Let \( f \) be \( \Lambda^{-1}1^+1^+ \); then the term \( S_0 \) above is in \( \text{Dom} f \) and \( (S_0)^f \) is precisely \( T_0 \) above; performing the algorithm leads to the normal form expression

\[
[\Lambda^{-1}][1^+][1^+] = [\Lambda^{(1)}][\Lambda^{(1)}][\Lambda^{(-3)}] \quad (\text{i.e., } [\Lambda^+][\Lambda^+][00^-][0^-][\Lambda^-]).
\]

5. The case of left distributivity

Let us finally consider now the variety \( \mathcal{L} \mathcal{D} \) formed by left distributive magmas. \( \mathcal{L} \mathcal{D} \) is defined by the equation

\[
x \times (y \times z) = (x \times y) \times (x \times z).
\]

This equation is balanced (and minimal), but its second member is obviously not linear, and things are in that case much more complicated. Nevertheless the approach developed in the preceding section works mutatis mutandis. In particular, the corresponding monoid \( \mathcal{L} \mathcal{D} \mathcal{D} \) is still associated with a group, and this group is a natural extension of the infinite braid group \( B_\infty \). The word problem in free left distributive magmas is connected with the existence of normal forms for the members of this group, and the approach sketched in the case of associativity still works. The interest in left distributive magmas has been recently renewed by their connection with the algebras formed by polynomials of elementary embeddings in set theory (see [La, De4]), and also with theory of knots and 3-manifolds (see [Br, FR, Jo]). Their theory is far from being finished, and many results are still conjectured, but it already seems certain that the monoid \( \mathcal{L} \mathcal{D} \mathcal{D} \) will be a key tool in the elaboration of this study (see [De1–De3]).

We shall not go further in this note. We just hope that the reader will be convinced that the general framework described above could be applied to various similar cases.

**References**


Département de Mathématiques, Université de Caen, 14032 Caen-cédex, France

E-mail address: dehornoy@frcaen51.bitnet